DECENTRALIZED ESTIMATION OF LINEAR GAUSSIAN SYSTEMS*

by

David A. Castañon†

ABSTRACT

In this paper, we propose a framework for the design of linear decentralized estimation schemes based on a team-theoretic approach. We view local estimates as "decisions" which affect the information received by other decision makers. Using results from team theory, we provide necessary conditions for optimality of the estimates. For fully decentralized structures, these conditions provide a complete closed-form solution of the estimation problem. The complexity of of the resulting estimation algorithms is studied as a function of the performance measure, and in the context of some simple examples.

†The author is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

Report Documentation Page

1. REPORT DATE **1981**	2. REPORT TYPE	3. DATES COVERED **00-00-1981 to 00-00-1981**

4. TITLE AND SUBTITLE **Decentralized Estimation of Linear Gaussian Systems**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Massachusetts Institute of Technology,Laboratory for Information and Decision Systems,77 Massachusetts Avenue,Cambridge,MA,02139-4307**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **29**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

1. INTRODUCTION

A standard problem in estimation theory consists of using a set of available information about a random variable to obtain an estimate of its value. When the criterion used in evaluating the estimate is the conditional variance of the estimate, the best estimator is given by the conditional mean. However, this formulation assumes that all of the available information is concentrated at a central location. In many areas of application, such as Command and Control systems and meteorology, the acquisition of data is characterized by sensors which are spatially and temporally distributed. Thus, there are nontrivial costs associated with the transfer of data to a central location for the purpose of estimation.

An approach to designing estimation algorithms for these areas of application is to preprocess some of the data at various local processing nodes, thereby reducing the communication load on the system. The result is an estimation scheme with a fixed structure (often hierarchical), and constraints on the available information at any one node. Figure 1 depicts a typical estimator structure.
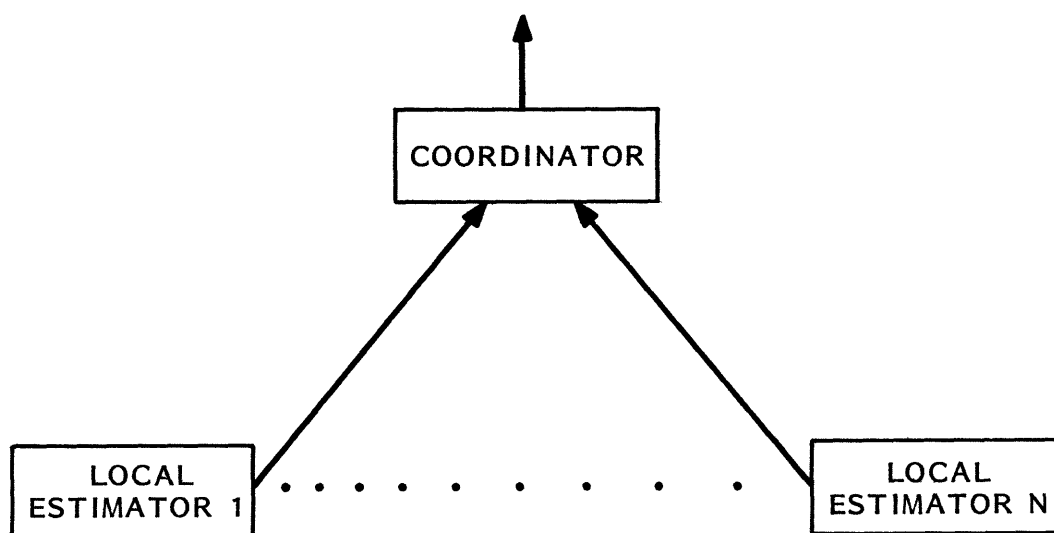

Figure 1

The structure of Figure 1 has similarities with a decentralized decision problem. In this paper, we propose to study estimation problems with fixed estimator structures, hereafter referred to as distributed estimation problems, by imbedding the estimation in a class of decentralized decision problems. These decision problems have special structures which can be exploited for some linear Gaussian systems to obtain closed-form solutions for the estimators. In particular, the decisions variables do not affect the evolution of the state variables and, in certain cases, they do not affect the observations received by other decision makers. This latter case results in a partially nested decision problem, as defined in Ho and Chu [1].

There has been a significant amount of recent work on the subject of distributed estimation. The various approaches can be divided into two classes: The first class consists of methods which use the distributed structure of the problem in such a way as to achieve an overall estimator whose error corresponds to that of a fully centralized estimator, and thus optimality is achieved. Elegant solutions to some of these problems are presented in [2], [3], and [4]. The second class of approaches consists of utilizing a fixed structure, which is simple, to achieve the best performance possible with this restricted structure. This approach can seldom achieve the performance of a centralized scheme. Typical of the results in this case are the papers of Tacker, Sanders and their colleagues [5], [6].

In this paper, we follow the spirit of the second approach. Specifically, we take as given a specific architecture of processing stations, with prespecified flows of information among them. Given this structure, and the apriori statistics of the random variables present in the system, we restrict the data processing to consist of linear strategies of the available data. It is our purpose to characterize the "best" processing schemes in terms of an overall performance measure; our estimation problem will thus become a stochastic team problem, where a number of decision agents with different information seek to minimize a common goal.

Fixed structure decentralized decision problems have been considered
by a number of authors [7], [8], and [9]. Our approach in this paper
follows very closely the formulation of Barta [9] for linear control of
decentralized stochastic systems. Indeed, most of the results of Section
4 of this paper appear in Barta and Sandell [10].

The paper is organized as follows. Section 2 contains the mathe-
matical formulation of fixed structure linear estimation problems using
a decision theoretic viewpoint. Section 3 presents general necessary
conditions which optimal estimators must statisfy. These conditions are
not very useful due to their complexity. In Section 4, we specialize
the results of Section 3 to a specific structure which corresponds to a
fully decentralized estimation algorithm. This case permits significant
analysis, as was previously done in Barta and Sandell []0]. We extend
their results to illustrate how the complexity of the local estimation
algorithm depends on the importance of correlation between the errors of
the various local estimators. Section 5 contains some simple examples
which illustrate the results of Section 4. Section 6 discusses the results
and areas of future research.

2. MATHEMATICAL FORMULATION

Assume that there are N local substations and one coordinator station
in the decentralized estimation systems. Denote the state of the environ-
ment by $x(t)$, an R^n-valued random process on $[0,T]$ whose evolution is
governed by the stochastic differential equation

$$dx(t) = A(t)x(t)dt + B(t)dw(t), \qquad (2.1)$$

where $w(t)$ is an R^m- valued standard Wiener process. Each local substation
receives data from local measurements, described by the observation equations

$$dy_i(t) = C_i(t)x(t)dt + D_i(t)dv_i(t) \qquad (2.2)$$

4

where $v_i(t)$, $w(t)$ are standard, mutually independent Wiener processes, and $y_i(t)$ is an R^{m_i} - valued random process. The matrices $A(t)$, $B(t)$, $C_i(t)$, $D_i(t)$ are assumed continuous an $[0,T]$ for $i = 0,...N$. In addition, the matrices $D_i(t)$ are assumed invertible for all i,t.

To each local substation corresponds a decision agent, whose decisions are denoted by $u_i(t)$ in R^{p_i}. The decisions made at each substation depend only on real-time observations of local data, as in equation (2.2), plus the apriori knowledge about the statistics of the systems. The apriori knowledge, common to all local substations and the coordinator station, consists of knowledge of the matrices $A(t)$, $B(t)$, $C_i(t)$, $D_i(t)$, for $i = 0,...N$, $t \varepsilon [0,T]$, together with the initial distribution of the initial condition $x(0)$. For the sake of simplicity, we assume that $x(0)$ is a zero-mean, normal random variable with covariance $\Sigma(0)$.

The coordinator station receives the decision outputs of all the local subsystems, $u_i(t)$, $i = 1,...N$, in addition to an independent set of measurements $y_0(t)$. The output of the coordinator station is denoted by $u_0(t)$, and it is based on real-time observation of measurements and the prior decisions of the local substations.

Associated with the estimation structure is a performance index, of the form

$$J = E\left\{ \int_0^T (\underline{u}(t) - S(t)x(t))^T Q(t) (\underline{u}(t) - S(t)x(t))dt, \right\} \qquad (2.3)$$

where $\underline{u}(t)$ consists of the vector of decisions,

$$\underline{u}^T(t) = (u_0^T(t),...,u_N^T(t)), \qquad (2.4)$$

and the superscript T denotes transposition. The matrix $Q(t)$ is assumed positive semidefinite and continuous for t in $[0,T]$. With this performance criterion, the design of a distributed estimation scheme can be reduced to determining the admissible decision strategies which minimize the quadratic function J.

The admissible strategies are restricted to be linear maps of the available information which yield mean-square integrable decision variables. Specifically, since equation (2.2) implies that the local observations are corrupted by additive white noise, we assume that, for i = 1,...n,

$$u_i(t) = \int_0^t H_i(t,s) dy_i(s) \qquad (2.5)$$

where

$$H_i(t,s) = 0 \quad \text{if } s > t, \qquad (2.6)$$

and

$$\text{Trace} \int_0^T \int_0^T H_i(t,s) H_i^T(t,s) dt ds < \infty. \qquad (2.7)$$

For the coordinator, we assume that

$$u_0(t) = \int_0^T H_0(t,s) dy_0(s) + \sum_{i=1}^N \int_0^t K_i(t,s) u_i(s) ds$$
$$+ \sum_{i=1}^n L_i(t) u_i(t) \qquad (2.8)$$

where H_0, K_i satisfy (2.6) and (2.7), while the matrices $L_i(t)$ are continuous on $[0,T]$.

The parametrization of the control laws in equations (2.5) to (2.8) results in admissible strategy spaces which are Hilbert spaces. Specifically, the admissible strategies for u_i, i = 1,...N, are elements of the Hilbert space of linear operators from $L_2([0,T], R^{n_i})$ to $L_2([0,T], R^{p_i})$ with finite trace, and inner product

$$\langle H^1, H^2 \rangle = \text{Trace} \int_0^T \int_0^T H^1(t,s) H^{2T}(t,s) dt ds = \text{Trace} (H_1 H_2^*). \qquad (2.9)$$

For additional information about Hilbert spaces of operators, the reader should consult Balakrishnan [11]. We will use the symbol H_i without its arguments to refer to the linear operator, while $H_i(t,s)$ will be used to refer to the kernel of the operator.

6

The assumption of linear strategies for all decision agents in the problem represents a restriction on the class of admissible strategies. However, the system and observations described by equations (2.1) and (2.2) result in zero-mean, jointly Gaussian random processes $x, y_o, \ldots y_N$. Since the decisions $\underline{u}(t)$ do not affect the evolution of the state $x(t)$ (this is a property of estimation problems) for any control law $\underline{u}(t)$ such that

$$E \int_O^T ||\underline{u}(t)||^2 dt < \infty, \tag{2.10}$$

we can use a version of Fubini's theorem to show

$$J = \int_O^T E \left\{ (\underline{u}(t) - S(t)x(t))^T Q(t)(\underline{u}(t) - S(t)x(t)) \right\} dt. \tag{2.11}$$

Notice that the optimal estimator will minimize the integrand

$$J_1 = E \left\{ (\underline{u}(t) - S(t)x(t))^T Q(t)(\underline{u}(t) - S(t)x(t)) \right\} \tag{2.12}$$

almost everywhere. In many cases, this will enable us to show that the true optimal solution belongs to the admissible class of linear strategies. To conclude this section, we will discuss some relevant examples, and indicate how they fit in this framework.

Example 1: Centralized estimation

Assume that $N = 0$, so that the only station present is the coordinator station. In this case, J_1 corresponds to

$$J_1 = E \left\{ (u_o(t) - S(t)x(t))^T Q(t)(u_o(t) - S(t)x(t)) \right\}.$$

Its minimum among all mean-square integrable $u_o(t)$ is achieved at

$$u_o(t) = S(t)\hat{x}(t) \tag{2.13}$$

where $\hat{x}(t)$ is the minimum variance estimate of x_t, given the prior observations, which is obtained from a Kalman filter. Hence, the optimal estimator is linear.

Example 2.   Hierarchical Estimation

Let N = 2. Furthermore, let $p_0 = p_1 = p_2 = n$ and

$$S(t) = \begin{bmatrix} I \\ I \\ I \end{bmatrix} \qquad Q(t) = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}.$$

Assume $C_0(t) \equiv 0$.

Then, equation (2.12) yields

$$J_1 = E \left\{ \sum_{i=0}^{2} (u_i(t) - x(t))^T (u_i(t) - x(t)) \right\}.$$

We consider the minimization of J_1 over all mean-square integrable decision. The last two terms in the sum are minimized by using local Kalman filters at each local substation. Furthermore, it was established in Willsky, Castanon et al [2], that the first term can be minimized absolutely, when the local strategies are Kalman filters, by a strategy of the form (2.8). Hence, the optimal hierarchical estimator for this problem is in the class of linear estimators.

Example 3.   Fully Decentralized Estimation

Assume that there is no coordinator station, so that $u_0(t) \equiv 0$ for all t. In this case,

$$J_1 = E \left\{ (\underline{u}(t) - s(t)x(t))^T Q(t) (\underline{u}(t) - s(t)x(t)) \right\}.$$

For each t, this is a static team problem with jointly Gaussian statistics; hence, Radner's theorem [12] implies that the optimal decision strategies are linear maps of the available observations, and hence they belong to the linear class in equations (2.5) to (2.8).

Example 4.   Let N = 1, $p_1 = 1$, $p_0 = n$, and

$$s(t) = \begin{bmatrix} I \\ 0 \end{bmatrix} , \qquad Q(t) = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

Then,

$$J_1 = E \left\{ (u_o(t) - x(t))^T (u_o(t) - x(t)) \right\}$$

It is clear that, if n > 1, some form of nonlinear encoding of the infor-
mation y_1 will provide a lower value of J_1 than the best linear encoder,
because u_1 is a scalar signal and x is a vector process. In this case, the
optimal decision rules are nonlinear.

In many cases, the optimal estimation strategies will be nonlinear.
Nevertheless, there will be a person-by-person-optimal linear strategy
which will be of interest because of ease of implementation. In the next
Section, we provide necessary conditions which characterize these linear
person-by-person optimal strategies.

3. NECESSARY CONDITIONS

The formulation of Section 2 imbedded the distributed estimation pro-
blem into a team decision problem with a quadratic criterion, where decision
rules are elements of a Hilbert space of linear operators. In this section,
we provide necessary conditions which characterize the estimators resulting
from this approach. The mathematical development of this section follows
closely the development in Barta [9].

In operator notation, equations (2.5) and (2.8) can be written as

$$u_i = H_i(dy_i), \qquad i = 1, \ldots N \qquad (3.1)$$

$$u_o = H_o dy_o + \sum_{1=1}^{N} (K_i u_i + L_i u_i) \qquad (3.2)$$

where L_i is the linear operator with kernel

$$L_i(t,s) = L_i(t)\delta(t-s) \qquad (3.3)$$

Furthermore, the quadratic functional (2.4) can be written as

$$J = E\left\{\int_O^T (\underline{u}(t)-S(t)x(t))^T Q(t)(\underline{u}(t)-S(t)x(t))dt\right\}$$

$$= \text{Trace}\left\{S*QS\textstyle\sum_{xx} + Q\textstyle\sum_{\underline{uu}} - 2Q\textstyle\sum_{\underline{ux}} \ S*\right\} \tag{3.4}$$

where $\sum_{xx}$, $\sum_{\underline{ux}}$, and $\sum_{\underline{uu}}$ are the covariance operators [11] corresponding to the random processes $x(t)$ and $\underline{u}(t)$. Note that the decision operators are implicit in defining $\underline{u}(t)$ as a random process.

Let's partition $\underline{u}$ as

$$\underline{u}(t) = [u_o(t) \mid u_1(t)...u_N(t)]^T = [u_o(t),\bar{u}(t)]^T \tag{3.5}$$

Then, $\sum_{\underline{uu}}$ can be partitioned

$$\sum_{\underline{uu}} = \begin{bmatrix} \sum_{u_o u_o} & \sum_{u_o \bar{u}} \\ \sum^*_{u_o \bar{u}} & \sum_{\bar{u}\bar{u}} \end{bmatrix} \tag{3.6}$$

Furthermore, $\bar{u}(t)$ is related to $\underline{y}(t)$ by

$$\bar{u}(t) = \begin{bmatrix} H_1 & & \\ & \ddots & \\ & & H_N \end{bmatrix} \underline{y}(t) \tag{3.7}$$

$$= \text{diag}\{H_i\} \ \underline{y}(t)$$

so that

$$\sum_{\bar{u}\bar{u}} = (\text{diag } H_i) \sum_{d\underline{y}d\underline{y}} (\text{diag } H_i)^* \tag{3.8}$$

Similarly,

$$\sum_{u_o \bar{u}} = \begin{bmatrix} H_o \sum_{dy_o dy_1} H_1^* \cdots H_o \sum_{dy_o dy_N} H_N^* \end{bmatrix} +$$

$$+ \sum_{i=1}^{N} [(K_i+L_i) H_i \Sigma_{dy_i dy_1} H_1^* \cdots (K_i+L_i) H_i \Sigma_{dy_i dy_N} H_N^*] \qquad (3.9)$$

and

$$\Sigma_{u_o u_o} = H_o \Sigma_{dy_o dy_o} H_o^* + \sum_{i=1}^{N} \{ H_o \Sigma_{dy_o dy_i} H_i^* (K_i^* + L_i^*)$$

$$+ (K_i+L_i) H_i \Sigma_{dy_o dy_i}^* H_o^* \}$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{N} (K_i+L_i) H_i \Sigma_{dy_i dy_j} H_j^* (K_j+L_j)^* \qquad (3.10)$$

A similar partition yields

$$\Sigma_{\underline{u}x} = \begin{bmatrix} \Sigma_{u_o x} \\[2ex] \Sigma_{\bar{u}x} \end{bmatrix} \qquad (3.11)$$

where

$$\Sigma_{u_o x} = H_o \Sigma_{dy_o x} + \sum_{i=1}^{N} (K_i+L_i) H_i \Sigma_{dy_i x} \qquad (3.12)$$

$$\Sigma_{\bar{u}x} = [H_i \Sigma_{dy_1 x} \cdots H_N \Sigma_{dy_N x}] \qquad (3.12)$$

Using equations (3.6) – (3.13) in equation (3.4), we can express the functional J as a deterministic quadratic function of the operators H_i, L_i, K_i, which are elements of a linear Hilbert space. We will denote this dependence by

$$J = J (\bar{H}, \bar{L}, \bar{K}) \qquad (3.14)$$

11

Since J is a quadratic functional, and the linear operators $\bar{H}$, $\bar{L}$, $\bar{K}$ are elements of Hilbert spaces, we can compute the Frechet differential of J [13] with respect to variations in the operators. In particular, we will denote the Frechet differential of J in the direction of each of the components of $\underline{H}$, $\underline{K}$ and $\underline{L}$. Partition the operators Q, S, according to equations (3.6), as

$$Q = \begin{pmatrix} Q_{oo} & Q_{ol} \\ Q_{lo} & Q_{ll} \end{pmatrix} \tag{3.14}$$

$$S = \begin{pmatrix} S_o \\ S_l \end{pmatrix} \tag{3.15}$$

Then, we can use equations (3.6) - (3.15) to obtain the Frechet differentials:

$$\delta_{H_o} J(\bar{H},\bar{K},\bar{L},\tilde{H}_o) = 2 \text{ Trace } \left\{ \left[Q_{oo} H_o \mathcal{L}_{dy_o dy_o} + \sum_{i=1}^{N} Q_{oo}(K_i+L_i) H_i \mathcal{L}^*_{dy_o dy_i} \right. \right.$$

$$+ Q_{ol} \begin{bmatrix} H_1 \mathcal{L}^*_{dy_o dy_1} \\ H_N \mathcal{L}^*_{dy_o dy_N} \end{bmatrix} - Q_{oo} S_o \mathcal{L}^*_{dy_o x}$$

$$\left. \left. - Q_{ol} S_l \mathcal{L}^*_{dy_o x} \right] \tilde{H}^*_o \right\} \tag{3.16}$$

$$\delta_{K_i+L_i} J(\bar{H},\bar{K},\bar{L};\tilde{K}_i+\tilde{L}_i) = 2 \text{ Trace } \left\{ \left[Q_{oo} H_o \mathcal{L}_{dy_o dy_i} H^*_i \right. \right.$$

$$+Q_{oo}\sum_{j=1}^{N} (K_j+L_j) H_j \mathcal{L}_{dy_j dy_i} H^*_i + Q_{ol} \begin{bmatrix} H_1 \mathcal{L}^*_{dy_i dy_1} H^*_i \\ \vdots \\ H_N \mathcal{L}^*_{dy_i dy_N} H^*_i \end{bmatrix}$$

$$\left. \left. - Q_{oo} S_o \mathcal{L}^*_{dy_i x} H^*_i - Q_{ol} S_l \mathcal{L}^*_{dy_i x} H^*_i \right](\tilde{K}^*_i+\tilde{L}^*_i) \right\} \tag{3.17}$$

12

$$\delta_{H_i} J(\bar{H},\bar{K},\bar{L} \; ; \; \tilde{H}_i) = 2 \; \text{Trace} \left\{ \left[(K_i^* + L_i^*) Q_{oo} H_o \int_{dy_o dy_i} \right. \right.$$

$$+ \sum_{j=1}^{N} (K_i + L_i)^* Q_{oo} (K_j + L_j) H_j \int_{dy_j dy_i}$$

$$+ Q_{lo}^i H_o \int_{dy_o dy_i} + (K_i + L_i)^* \begin{bmatrix} Q_{ol}^i H_1 \int_{dy_i dy_1}^* \\ \\ Q_{ol}^N H_N \int_{dy_i dy_N}^* \end{bmatrix}$$

$$+ Q_{lo}^i \sum_{j=1}^{N} (K_j + L_j) H_j \int_{dy_j dy_i} + \sum_{j=1}^{N} Q_{11}^{ij} H_j \int_{dy_j dy_i}$$

$$- (K_i + L_i)^* (Q_{oo} S_o + Q_{ol} S_1) \int_{dy_i x}^*$$

$$\left. \left. - (Q_{lo} S_o + Q_{11} S_1)^i \int_{dy_i x}^* \right] \tilde{H}_i^* \right\} \tag{3.18}$$

where Q_{lo}^i, Q_{11}^{ii}, S^i are the blocks partition in the corresponding partition of $\bar{u}(t) = (u_,(t), \ldots u_N(t))^T$.

Using expressions (3.16) - (3.18), we can provide necessary conditions for optimality of a set of linear maps $(\bar{H},\bar{K},\bar{L})$, as follows:

<u>Proposition 3.1</u>   If $\bar{H},\bar{K},\bar{L}$ minimize the functional J over the space of all linear maps, then

(a) $\delta J_{H_o}(\bar{H},\bar{K},\bar{L} \; ; \; \tilde{H}_o) = 0$

$\delta J_{K_i + L_i}(\bar{H},\bar{K},\bar{L} \; ; \; \tilde{K}_i + \tilde{L}_i) = 0$

$\delta J_{H_i}(\bar{H},\bar{K},\bar{L} \; ; \; \tilde{H}_i) = 0$

13

for all $i=1,\ldots N$, and for all admissible $\tilde{K}_i, \tilde{L}_i, \tilde{H}_i$ and $\tilde{H}_o$.

<u>Proof</u> The proof follows directly from Theorem 1 in Chapter 7 in Luenberger [13], since the existence of Frechet differentials provides an expression for the Gateaux differential, which must be zero at a minimum.

Proposition 3.1 can be used, together with the fact that admissible operators $\bar{H}, \bar{K}, \bar{L}$ are Volterra integral operators, to obtain sets of coupled integral conditions which characterize the optimal solution, in a manner similar to Wiener-Hopf factorization [14]. We will not do so here, focusing instead on obtaining the expressions which characterize the optimum in the specific case of equations (2.2) - (2.3) for the fully decentralized case in the next section.

4. FULLY DECENTRALIZED ESTIMATION

In the fully decentralized case, the coordinator station is absent. In terms of the formulation of section 3, the operators K_i, L_i and H_o are identically zero, as are the weighting matrices S_o, Q_{oo}, Q_{ol} and Q_{lo}, for all time t in $[0,T]$. This causes an extensive simplification in the equations of Proposition 3.1. Specifically, equation (3.18) now becomes

$$\delta_{\tilde{H}_i} J(\bar{H}; \tilde{H}_i) = \text{Trace} \left\{ \left[\sum_{j=1}^{N} 2Q_{11}^{ij} H_j \, \Sigma_{dy_j dy_i} - (Q_{11}S_1)^i \Sigma_{dy_i x}^* \right] \tilde{H}_i^* \right\} \quad (4.1)$$

The equivalent set of integral equations corresponding to equation (4.1) are

$$\sum_{j=1}^{N} Q_{11}^{ij} \int_0^t H_j(t,s_1) \Sigma_{dy_j dy_i}(s_1,s) ds_1 = (Q_{11}S_1(t))^i \Sigma_{xdy_i}(t,s) \quad (4.2)$$

A similar equation can be found in Barta-Sandell [10], where a solution is found using an innovations approach. We will present a different derivation of their results in this section.

Assume that $Q_{11} > 0$ and is constant in time. This implies that the cost functional J is strictly convex, so that there is a unique minimum, which is characterized by the integral equation (4.2). Furthermore, assume, without loss of generality, that all decisions u_i are scalar-valued, that is $p_i = 1$ for all i. A vector-valued decision can be decomposed into p_i stations with the same information. Hence, the assumption in equation (2.3) that the v_i are mutually independent Wiener processes will be removed at this stage, to allow for this development.

We begin by noting that equation (4.2) is a linear equation driven by a sum of terms in the right hand side. Hence, by superposition, the optimal solution $\hat{H}_j(t,s)$ can be written as

$$\hat{H}_j(t,s) = \sum_{\ell=1}^{N} \sum_{k=1}^{n} G_j^{\ell k}(t,s) \, S_{\ell k}(t) \tag{4.3}$$

where $G_j^{\ell k}(t,s)$ minimizes J when $S = \delta_{\ell k}$, that is, it has a one in the ℓk th entry and zero elsewhere. Hence, $G_j^{\ell k}(t,s)$ solves

$$\sum_{j=1}^{N} Q_{11}^{ij} \int_0^t G_j^{\ell k}(t,s_1) \sum_{dy_j dy_i} (s_1,s) ds_1 = Q_{11}^{i\ell} \sum_{x_k dy_i} (t,s) \tag{4.4}$$

Notice that the form of Q determines the form of the linear system on the left side. It is possible to solve for all $G_j^{\ell k}$ simultaneously, because of the consistency of the problems (4.4). Let $J^{\ell k}$ denote the cost function J when $s = \delta_{\ell k}$. Then,

$$(G_1^{\ell k}, \ldots G_n^{\ell k}) = \underset{\bar{G}^{\ell k}}{\text{argmin}} \, J^{\ell k}(\bar{G}^{\ell k}) \tag{4.5}$$

Define a global cost J^T, given by

$$J^T(\bar{G}^{1,1}, \ldots \bar{G}^{N,n}) = \sum_{\ell=1}^{N} \sum_{k=1}^{N} J^{\ell k}(\bar{G}^{\ell k}) \tag{4.6}$$

The cost J^T is separable in its arguments. Hence, minimization of J^T corresponds to solving equation (4.5) for each ℓ,k.

Let's examine closely the nature of the costs $J^{\ell k}$. From equation (2.4), $J^{\ell k}$ corresponds to

$$J^{\ell k} = E \left\{ \int_o^T (\underline{u}(t) - \delta_{i-\ell} x_k(t))^T Q (\underline{u}(t) - \delta_{i-\ell} x_k(t)) dt \right\} \qquad (4.7)$$

where $\delta_{i-\ell}$ is a vector with all zeroes except a one in the ℓ'th entry. Furthermore, minimization of $J_\ell^{\ell k}$ is accomplished by minimizing

$$J_1^{\ell k} = E \quad (\underline{u}(t) - \delta_{i-\ell} x_k(t))^T Q(\underline{u}(t) - \delta_{i-\ell} x_k(t)) \qquad (4.8)$$

for each t. Let $d_i(t)$ correspond to the n x N matrix

$$d_i(t) = \begin{pmatrix} \underline{u}^{i1} \\ \vdots \\ \underline{u}^{in} \end{pmatrix} \qquad (4.9)$$

representing the decision variables associated with problems J_1^{ik}, k=1,...n in (4.6). Let D(t) be

$$D(t) \equiv \begin{bmatrix} d_1(t) \\ \vdots \\ d_N(t) \end{bmatrix}$$

Let

$$X(t) = \begin{bmatrix} x(t) & & & \\ & x(t) & & \\ & & \ddots & \\ & & & X(t) \end{bmatrix}$$

be an n N x N matrix. Then, a simple calculation establishes that

$$J_1^T = \text{Trace } E \left\{ (D(t)-X(t)) \ Q(D(t)-X(t))^T \right\} \qquad (4.10)$$

where the i-th column of D(t) is a linear function of the local observation process $y_i(t)$ only.

16

This is the same formulation used in Barta-Sandell [10]. We will state their main result without proof, as it applies to systems of the form (2.2) - (2.4). Before we can do so, we must introduce some notation.

The state process of equation (2.2) is given by

$$dx(t) = A(t) x(t)dt + B(t)dw(t) \qquad (4.11)$$

with local observations

$$dy_i(t) = C_i(t) x(t)dt + D_i(t)dv_i(t) \qquad (4.12)$$

where $v_i(t)$, $w(t)$ are standard Brownian motions with $w(t)$ independent of all $v_i(s)$.

Let

$$\mathcal{A}(t) = \text{diag} \left\{ A(t), \ldots A(t) \right\}$$

$$\mathcal{B}(t)dw(t) = \text{diag} \left\{ B(t)dw(t), \ldots B(t)dw(t) \right\}$$

$$C(t) = \text{diag} \left\{ C_1(t), \ldots C_n(t) \right\} \qquad (4.13)$$

then, we have

$$dX(t) = \mathcal{A}(t)X(t)dt + \mathcal{B}(t)dw(t) \qquad (4.14)$$

Define also

$$\Sigma_{ww}(t) = \begin{bmatrix} \mathcal{Q}_{11}I & \cdots & \mathcal{Q}_{1N}I \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \mathcal{Q}_{N1}I & \cdots & \mathcal{Q}_{NN}I \end{bmatrix} \cdot \text{diag} [B(t)B^T(t), \ldots B(t)B^T(t)] \qquad (4.15)$$

as the enlarged system relevant driving noise intensity.

Similarly, define

$$\Sigma_{vv}(t) = \begin{bmatrix} Q_{11}D_1(t)D_1^T(t) & \cdots & Q_{1N}E\left\{D_1(t)dv_1(t)dv_N(t)D_N^T(t)\right\} \\ & & \\ Q_{N1}E\left\{D_N(t)dv_N(t) \quad dv_1(t)D_N^T(t)\right\} & \cdots & Q_{NN}D_N(t)D_N^T(t) \end{bmatrix} \qquad (4.16)$$

as the enlarged system <u>relevant</u> observation noise intensity. With this notation, the main result of [10] is:

<u>Proposition 4.2</u> <u>The Decentralized Kalman Filter</u>

The optimal team decision rule for equation (4.10), $\hat{X}(t)$, satisfies

$$d\hat{x}_i(t) = A(t)\hat{x}_i(t)dt + K(t)[I_i dy_i(t) - C(t)\hat{x}_i(t)] \qquad (4.17)$$

where

$$K(t) = \Sigma(t)C'(t)\Sigma_{vv}^{-1}(t)$$

$I_i = [0^T, \ldots I^T, \ldots 0^T]^T$ is a $\sum\limits_{j=1}^{M} m_j \times m_i$ dimensioned matrix with the identity in its ith block, and $\Sigma(t)$ solves the Ricatti equation

$$\Sigma = A(t)\Sigma + \Sigma A^T(t) - K(t)\Sigma_{vv}K^T(t) + \Sigma_{ww} \qquad (4.16)$$

$$\Sigma(o) = \begin{bmatrix} Q_{11}I & \cdots & Q_{1N}I \\ \vdots & & \\ Q_{N1}I & \cdots & Q_{NN}I \end{bmatrix} \quad \text{diag } [\Sigma_o, \ldots \Sigma_o].$$

The estimator of Proposition 4.2 is depicted in Figure 2. The striking feature of this estimator is that each local agent uses identical estimation systems, of dimension NnxN, differing only in the input used to drive the systems. However, in many applications, these estimators are much larger than are necessary. In particular, it is important to note that it is the presence of Q which creates nontrivial couplings in the team problem, leading to large-dimension estimators.
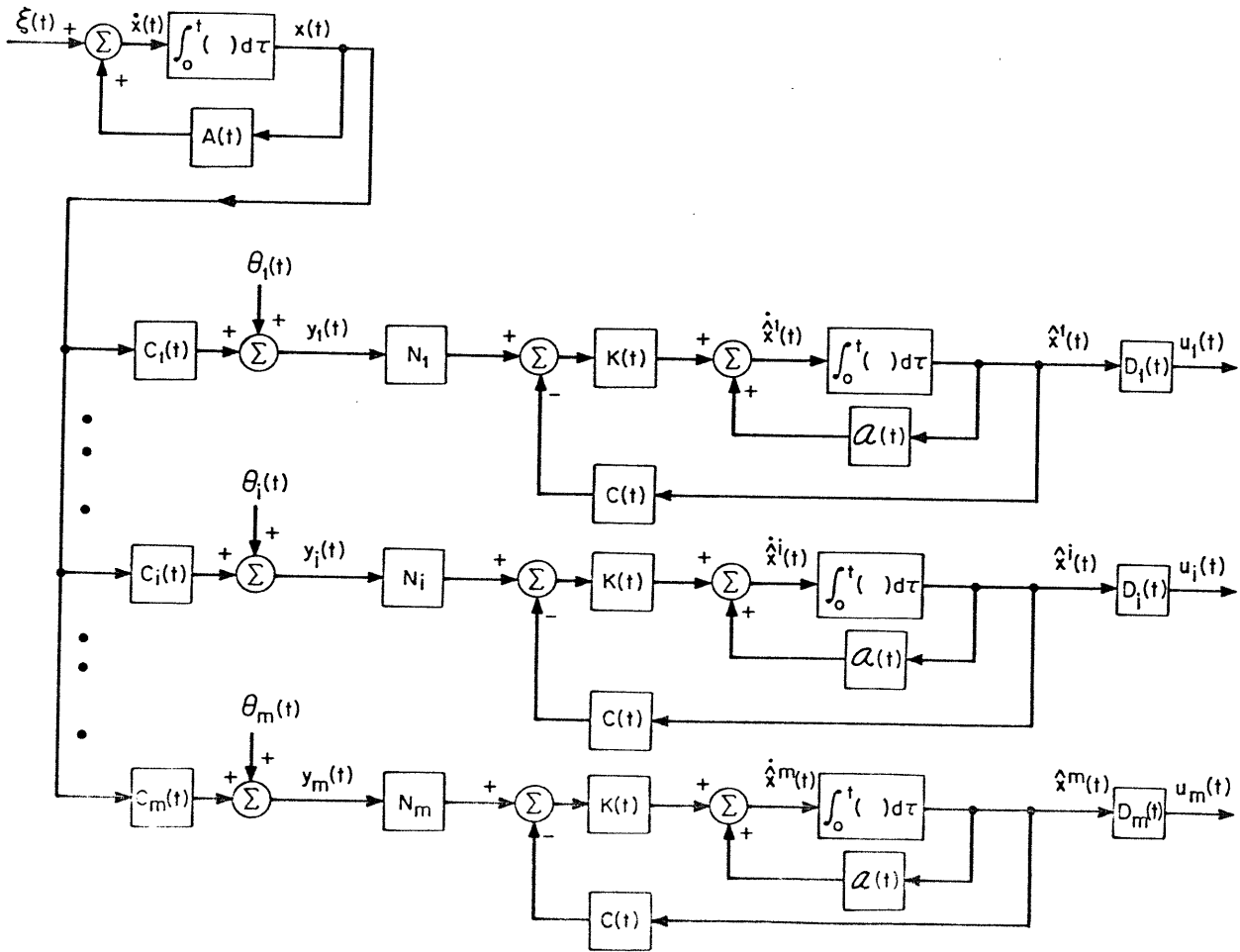
Figure 2

When Q is diagonal, the expresions for $\sum_{ww}(t)$ and $\sum_{vv}(t)$ are block-diagonal. In this case, it can be established that $\sum(t)$, as given by equation (4.16) will also be block-diagonal, and the optimal estimator will decompose into blocks of much smaller dimension. We formalize this in the following proposition.

Proposition 4.3 Assume Q is diagonal. Then, the optimal decision rule which minimizes (4.10) can be synthesized using n-dimensional estimators at each local station.

The proof follows directly from equations (4.15) and (4.16). In the next section, we will study some specific examples to illustrate the comlexity of the algorithm of Proposition 4.2, and the relation of the off-diagonal elements of the matrix Q with this complexity.

5. EXAMPLES

In this section, we discuss some examples of fully decentralized estimation problems, indicating their relation with the results of section 4. To facilitate the understanding of the examples, we will discuss only non-dynamic Gaussian systems.

Example 1. Let x_1, x_2 be independent, zero-mean Gaussian random variables with unit variance. Define the two observation equations

$$y_1 = x_1 + v_1 \qquad (5.1)$$

$$y_2 = x_2 + v_2 \qquad (5.2)$$

where v_1, v_2, x_1, x_2 are mutually independent, normal, zero-mean random variables with unit variance.

Assume that there are two local substations. Each substation i has access to its own measurement y_i. The performance of the elements is to be evaluated as

$$J = E \left\{ \left[\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right]^T \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} \left[\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right] \right\}$$

$$= E \left\{ (u_1 - x_1 - x_2)^2 + (u_2 - x_2)^2 + (u_1 - x_1 - x_2)(u_2 - x_2) \right\}. \tag{5.3}$$

Conditioning on y_1 inside the expectation of equation (5.3), and differentiating with respect to u_1 yields

$$2u_1 - 2 E \{x_1 | y_1\} = 0$$

Similarly, conditioning on y_2 and differentiating with respect to u_2 yields

$$2u_2 - 2E \{x_2 | y_2\} - E \{x_2 | y_2\} = 0$$

Hence,

$$u_1 = E \{x_1 | y_1\} \tag{5.4}$$

$$u_2 = \frac{3}{2} E \{x_2 | y_2\}$$

In this example, $S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. If $S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, it is clear that

$$u_1 = E \{x_1 | y_1\} \tag{5.5}$$

$$u_2 = E \{x_2 | y_2\}$$

is the optimal decentralized estimator. Now, let $S = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

Then,

$$J = E \{ (u_1 - x_2)^2 + u_2^2 + (u_1 - x_2)u_2 \}$$

conditioning with respect to y_1 and differentiating with respect to u_1 yields

21

$$2u_1 = 0$$

Repeating for y_2 and u_2 yields

$$2u_2 - E\{x_2|y_2\} = 0$$

Hence, for $S = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, the optimal strategy is

$$u_1 = 0$$

$$u_2 = 1/2 \ E\{x_2|y_2\}.$$

As indicated in Section 4, the solution for $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is the superposition of the solutions for $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $S = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

The presence of the off diagonal elements of $Q = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$ is important in creating the nature of the solution. Notice that, in spite of the independence x_1, y_1 and x_2, y_2, that the optimal estimator for $S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is not

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} E\{x_1|y_1\} \\ E\{x_2|y_2\} \end{bmatrix}$$

Example 2

Assume, in example 1, that $x_1 = x_2$. Repeating the same logic for obtaining the optimal solution for $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, we obtain the sufficient conditions

$$2u_1 - 5 \ E\{x_1|y_1\} + E\{u_2|y_1\} = 0$$

$$2u_2 - 4 \ E\{x_2|y_2\} + E\{u_1|y_2\} = 0$$

$$(5.5)$$

the coupled equations (5.5) can be solved by noting that $u_1 = ay_1$, $u_2 = by_2$, for some constant a, b. Equation (5.5) becomes

$$2u_1 - 5\ E\{x_1|y_1\} + bE\{y_2|y_1\} = 0$$

$$\text{(5.6)}$$

$$2u_2 - 4\ E\{x_2|y_2\} + aE\{y_1|y_2\} = 0$$

Now,

$$E\{y_2|y_1\} = E\{x_1|y_1\}$$

$$E\{y_1|y_2\} = E\{x_2|y_2\}$$

So,

$$a\ y_1 = \left(\frac{5}{2} - \frac{b}{2}\right) E\{x_1|y_1\}$$

$$\text{(5.7)}$$

$$b\ y_2 = 2 - a/2\ E\{x_2|y_2\}$$

Rewriting in terms of contants,

$$a + b/4 = 5/4$$

$$b + a/4 = 1$$

so $\qquad$ $a = 16/15$ $\qquad\qquad\qquad\qquad\qquad$ (5.8)

$\qquad\qquad$ $b = 11/15$

Equation (5.8) was obtained by solving the simultaneous euqations obtained from the variational arguments. For differential systems, these equations will be coupled integral equations which are hard to solve.

Let's establish the solution (5.8) using the decomposition approach of Section 4. Let $S = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Then, the performance measure is

$$J = E\{(u_1-x_1)^2 + (u_1-x_1)u_2+u_2^2\}$$

Variational arguments yield

$$2u_1 - 2 E\{x_1|y_1\} + E\{u_2|y_1\} = 0$$

$$2u_2 - E\{x_2|y_2\} + E\{u_1|y_2\} = 0 \qquad\qquad (5.9)$$

which imply $a = 7/15$, $b = 2/15$

By symmetry, the solution for $S = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ is

$$a = 2/15 \qquad b = 7/15$$

For $S = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, the performance measure is

$$J = E\{(u_1-x_2)^2 + (u_1-x_2)u_2 + u_2^2\}$$

$$= E\{(u_1-x_1)^2 + (u_1-x_1)u_2 + u_2^2\}$$

which has already been solved, yielding

$$a = 7/15 \qquad b = 2/15$$

Summary all three yields the result for $S = \begin{matrix} 1 & 1 \\ 0 & 1 \end{matrix}$ as

$$a = 7/15 + 2/15 + 7/15 = 16/15$$

$$b = 4/15 + 7/15 = 11/15$$

We will now use proposition 4.2 directly to solve example 2. Since $x_1 = x_2$, the effective state dimension is 1. Hence, the matrix D in Section 4 has dimension 2 x 2, with the first column a function of y_1, while the second column is a fucntion of y_2. The overall team cost is given as in (4.10), by

$$J = \text{Trace} \left[(D - \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}) \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} (D- \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix})^T \right]$$

The optimal solution $\hat{X}$ is characterized by

$$E\left\{\left(\hat{X} - \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}\right) Q\, D^T\right\} = 0 \tag{5.10}$$

for any D whose first column is a function of y_1, and its second column a function of y_2. Let

$$\hat{X} = \begin{pmatrix} a_1 y_1 & b_1 y_2 \\ a_2 y_1 & b_2 y_2 \end{pmatrix} \tag{5.11}$$

Equations (5.10) and (5.11) imply

$$E\left\{\begin{pmatrix} a_1 y_1 - x & b_1 y_2 \\ a_2 y_1 & a_2 y_2 - x \end{pmatrix} Q \begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix}\right\} = 0 \tag{5.12}$$

which reduces to

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} E\left\{\begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix} Q \begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix}\right\} = E\left\{\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} Q \begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix}\right\} \tag{5.13}$$

Let's compute the terms in equations (5.13).

$$E\left\{\begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix}\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}\begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix}\right\} = \begin{pmatrix} 2 & 1/2 \\ 1/2 & 2 \end{pmatrix}$$

$$E\left\{\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} Q \begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix}\right\} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$$

Hence,

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1/2 \\ 1/2 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 7/15 & 2/15 \\ 2/15 & 7/15 \end{pmatrix}$$

25

The solution for $S = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is thus

$$u_1 = (2 \cdot 7/15 + 2/15) \; y_1 = 16/15 \; y_1$$

$$u_2 = (2 \cdot 2/15 + 7/15) \; y_2 = 11/15 \; y_2,$$

as was established before.

Notice that a diagonal Q would have decoupled the problem by permitting a trivial inversion of a diagonal matrix, as predicted in proposition 4.3.

6. CONCLUSION

We have presented a framework for the design of distributed estimation schemes with specific architectures, based on a decision theoretic approach. For a fully decentralized architecture, explicit solutions to the estimation problem were described and illustrated with several examples. The examples illustrate that the complexity of the decentralized estimation scheme is critically dependent on the importance of the cross-correlation of errors in the local estimators, which are represented by the off-diagonal elements of the positive definite matrix Q. Most practical systems will want to weigh heavily the correlation of local errors. For example, in a distributed surveillance network, it is important that errors in location or detection at one local substation be corrected by other substations. In other words, it is very costly for all substations to err in the same way. This is reflected in the performance measure by the off-diagonal elements of Q.

The examples in Section 5 illustrate the high dimensionality required by the local estimators in order to compensate for correlations in their errors. It is our conjecture that the dimensionality of the local estimators is directly related to the number of off-diagonal elemets of Q.

When there is a coordinator station present, the results presented in Section 3 provide necessary conditions for the optimality of the estimation operators. Unfortunately, the coupling between decisions at the local sub-stations and the information available to the coordinator makes the analysis a difficult problem. We expect that, under some simplifying assumptions, the necessary conditions of Section 3 can lead to a solution, as in Section 4. Such results have been reported in Willsky, Castanon et al [2] for a simple class of performance measures.

The formulation of Section 2 can be extended to incorporate communication restrictions, as well as delays in the transmission of local decisions. These are areas which will be studied in the future.

REFERENCES

[1] Y. C. Ho, K. C. Chu, " Equivalence of Information Structure in Static
     and Dynamic Team Problems," IEEE Trans. on Auto. Control, Vol., AC-18,
     1973.

[2] A. S. Willsky, M. Bello, D. Castanon, B. C. Levy, G. Verghese,
     "Combining and Updating of Local Estimates and Regional Maps Along
     Sets of One-Dimensional Tracks," to appear in IEEE Trans. on Auto.
     Control.

[3] J. L. Speyer, "Computation and Transmission Requirements for a
     Decentralized L Q G Control Problem," IEEE Trans. Auto.Control,
     Vol. AC-24, No. 2, April 1979.

[4] B. C. Levy, D. A. Castanon, G. C. Verghese, A. S. Willsky, " A
      Scaterring Framework for Decentralized Estimation Problems,"
     MIT/LIDS paper 1075, March 1981, Submitted to Automatica.

[5] E. C. Tacker and C. W. Sanders, "Decentralized Structures for State
     Estimation in Large Scale Systems," Large Scale Systems, Vol. 1,
     No. 1, February 1980.

[6] C. W. Sanders, E. C. Tacker, T. D. Litton, R. Y. S. Ling, "Specific
     Structures for Large Scale Eatimation Algorithms Having Information
     Exchange," IEEE Trans. Auto Control, AC-23, No. 2, 1978.

[7] C. Y. Chong and M. Athans, "On the Periodic Coordination of Linear
     Stochastic Systems," Automatica, Vol. 12, July 1976.

[8] D. P. Looze and N.R. Sandell, "Decomposition of Linear Decentralized
     Stochastic Control Problems," ESL-P-288 Proc. 1977 JACC, San Francisco,
     California June 1977.

[9] S. M. Barta, On Linear Control of Decentralized Stochastic Systems,
     Ph.D. Thesis, MIT, July 1978.

[10] S. M. Barta and N. R. Sandell, "Certainty Equivalent Solutions of
     Quadratic Team Problems by a Decentralized Innovations Approach,"
     ESL P-799, MIT, Cambridge, MA, February 1978, Submitted to IEEE
     Trans. Auto Control.

[11] A. V. Balakrishnan, Applied Functional Analysis, Springer-Verlag,
     New York 1976.

[12] R. Radner, "Team Decision Problems," Annals of Mathematical Statistics,
     Vol. 33, 1962.

[13] D. G. Luenberger, Optimization by Vector Spare Methods, John Wiley,
     New York, 1968.

[14] H. Van Trees, <u>Detection, Estimation and Modulation Theory</u>, John Wiley, 1968.